AMENDMENT TO THE CLAIMS

1. (Previously presented) A computer readable storage media storing instructions readable by a computer which, when implemented, cause the computer to resolve an overlapping ambiguity string in an input sentence of an unsegmented language by performing steps comprising:
   - segmenting the sentence into two possible segmentations;
   - recognizing the overlapping ambiguity string in the input sentence as a function of the two segmentations;
   - obtaining probability information based on at least one context feature adjacent the overlapping ambiguity string; and
   - outputting an indication for selecting one of the two segmentations as a function of the obtained probability information.

2. (Previously presented) The computer readable storage media of claim 1, wherein obtaining the probability information comprises obtaining probability information from a language model based on the at least one context feature and a left or right portion of the overlapping ambiguity string.

3. (Previously presented) The computer readable storage media of claim 2 wherein the language model comprises a trigram model.

4. (Previously presented) The computer readable storage media of claim 2 wherein outputting an indication for selecting one of the two segmentations comprises classifying the probability information.

5. (Previously presented) The computer readable storage media of claim 4 wherein classifying comprises classifying using Naïve Bayesian Classification.

6. (Previously presented) The computer readable storage media of claim 1 wherein segmenting the sentence comprises performing a Forward Maximum Matching (FMM) segmentation of the input sentence and a Backward Maximum Matching (BMM) segmentation of the input sentence.

7. (Previously presented) The computer readable storage media of claim 6 wherein recognizing the overlapping ambiguity string comprises recognizing a segmentation O_f of the overlapping ambiguity string from the FMM segmentation and a segmentation O_b of the overlapping ambiguity string from the BMM segmentation.

8. (Previously presented) The computer readable storage media of claim 7 wherein selecting one of the two segmentations is a function of a set of context features associated with the overlapping ambiguity string.

9. (Previously presented) The computer readable storage media of claim 8 wherein the set of context features comprises words around the overlapping ambiguity string.

10. (Previously presented) The computer readable storage media of claim 8 wherein selecting one of the two segmentations comprises classifying the probability information of the set of context features and O_f.

11. (Previously presented) The computer readable storage media

of claim 10 wherein selecting one of the two segmentations comprises classifying the probability information of the set of context features and O_b.

12. (Previously presented) The computer readable storage media of claim 8 wherein selecting comprising determining which of O_f or O_b has a higher probability as a function of the set of context features.

13. (Previously presented) The computer readable storage media of claim 1 wherein the unsegmented language is Chinese.

14. (Currently amended) A method of segmentation of a sentence of an unsegmented language, the sentence having an overlapping ambiguity string (OAS), the method comprising:

- generating a Forward Maximum Matching (FMM) segmentation of the sentence;
- generating a Backward Maximum Matching (BMM) segmentation of the sentence;
- obtaining probability information based on at least one context feature and at least part of the recognized OAS for each of the FMM and BMM;; and
- outputting an indication for selecting one of the FMM segmentation and the BMM segmentation as a function of obtained probability information.

15. (Previously presented) The method of claim 14 wherein outputting includes selecting one of the FMM segmentation of the overlapping ambiguity string and the BMM segmentation of the overlapping ambiguity string based on higher probability.

16. (Previously presented) The method of claim 15 wherein obtaining probability information comprises using an N-gram model.

17. (Previously presented) The method of claim 16 wherein obtaining probability information comprises obtaining probability information about a first word of the overlapping ambiguity string.

18. (Previously presented) The method of claim 17 wherein obtaining probability information comprises using probability information about a last word of the overlapping ambiguity string.

19. (Previously presented) The method of claim 16 wherein obtaining probability information comprises using the N-gram model comprises using information about context words around the overlapping ambiguity string.

20. (Previously presented) The method of claim 16 wherein using the N-gram model comprises using information about a string of words comprising a first word of the overlapping ambiguity string and two context words to the left of the first word.

21. (Previously presented) The method of claim 20 wherein using the N-gram model comprises using information about a string of words comprising a last word of the overlapping ambiguity string and two context words to the right of the last word.

22. (Previously presented) The method of claim 15 wherein outputting includes using Naïve Bayesian Classifiers.

23. (Original) The method of claim 14 and further comprising receiving information from a lexical knowledge base comprising a trigram model.

24. (Original) The method of claim 23 and further comprising receiving an ensemble of Naïve Bayesian Classifiers.

25. (Currently amended) A method of constructing information to resolve overlapping ambiguity strings in an unsegmented language comprising:

recognizing overlapping ambiguity strings in a training data;

replacing the overlapping ambiguity strings with tokens; and

generating an N-gram language model comprising information on constituent words of the overlapping ambiguity strings and context features surrounding the overlapping ambiguity strings.

26. (Original) The method of claim 25 wherein generating the N-gram language model comprises generating a trigram model.

27. (Original) The method of claim 25 and further comprising generating an ensemble of classifiers as a function of the N-gram model.

28. (Original) The method of claim 25 wherein recognizing the overlapping ambiguity strings comprises:

generating a Forward Maximum Matching (FMM) segmentation of each sentence in the training data;

generating a Backward Maximum Matching

(BMM) segmentation of each sentence in the training data;

recognizing an OAS as a function of the FMM and the BMM segmentations of each sentence in the training data.

29. (Original) The method of claim 28 and further comprising generating an ensemble of classifiers as a function of the N-gram model.

30. (Previously presented) The method of claim 29 wherein generating the ensemble of classifiers includes approximating probabilities of the FMM and BMM segmentations of each overlapping ambiguity string as being equal to the product of individual unigram probabilities of individual words in the FMM and BMM segmentations respectively, of the overlapping ambiguity string.

31. (Previously presented) The method of claim 30 wherein generating the ensemble of classifiers includes approximating a joint probability of a set of context features conditioned on an existence of one of the segmentations of each overlapping ambiguity string as a function of a corresponding probability of a leftmost and a rightmost word of the corresponding overlapping ambiguity string.